

DAFTAR ISI

LEMBAR PENGESAHAN	i
LEMBAR PERNYATAAN ORISINALITAS	ii
ABSTRAK	iii
ABSTRACT	iv
KATA PENGANTAR	v
DAFTAR ISI	vi
DAFTAR GAMBAR DAN ILUSTRASI	viii
DAFTAR TABEL	x
DAFTAR ISTILAH	xii
DAFTAR LAMPIRAN	xiii
Bab I PENDAHULUAN	1
I.1 Latar Belakang	1
I.2 Perumusan Masalah	4
I.3 Tujuan Penelitian	4
I.4 Batasan Penelitian	4
I.5 Manfaat Penelitian	5
I.6 Sistematika Pelaporan	5
Bab II LANDASAN TEORI	6
II.1 Data <i>Preprocess</i>	6
II.2 Data <i>Profiling</i>	6
II.3 <i>Multi-Column Analysis: Outliers and Deduplication</i>	6
II.4 Variasi Arsitektur Teknologi Data Profiling menggunakan Open Source Platform	7
II.5 Pemrosesan Mekanisme Data <i>Profiling</i>	9
II.6 Pemilihan Metode Penelitian	9
Bab III METODOLOGI PENELITIAN	11
III.1 Konseptual Model	11
III.2 Sistematika Penelitian	12
III.2.1 Fase Inisiasi	12
III.2.2 Fase Proses Data <i>Profiling</i>	12
III.2.3 Fase Pengujian	13

Bab IV	ANALISIS DAN DESAIN.....	14
IV.1	Analisis Pemilihan Algoritma.....	14
IV.1.1	Analisis Pemilihan Algoritma Outliers.....	14
IV.1.2	Analisis Pemilihan Model untuk Weka Scoring.....	16
IV.1.3	Analisis Pemilihan Algoritma Deduplication	18
IV.2	Analisis Pemilihan <i>Open Source Tools</i>	20
Bab V	IMPLEMENTASI DAN PENGUJIAN.....	22
V.1	Implementasi Ruang Lingkup Studi Kasus.....	22
V.2	Implementasi logika menggunakan Pentaho Data Integration.....	23
V.2.1	Pembuatan Koneksi Database pada Pentaho Data Integration (Kettle)	23
V.2.2	Konfigurasi Kolom Variabel pada Pentaho Data Integration	24
V.2.3	Implementasi Logika <i>Outliers</i>	25
V.2.4	Implementasi Logika <i>Deduplication</i>	33
V.3	Pengujian	49
Bab VI	KESIMPULAN DAN SARAN	53
VI.1	Kesimpulan	53
VI.2	Saran	53
	DAFTAR PUSTAKA	54